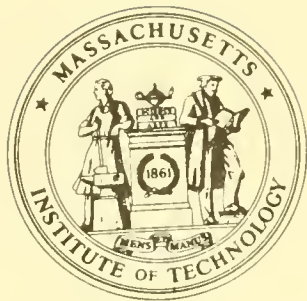






LIBRARY  
OF THE  
MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY





WORKING PAPER  
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

Constructive Duality for Discrete Optimization<sup>†</sup>

by

M. L. Fisher<sup>\*</sup> and J. F. Shapiro<sup>\*\*</sup>

WP 625-72

November, 1972

Preliminary Version

MASSACHUSETTS  
INSTITUTE OF TECHNOLOGY  
50 MEMORIAL DRIVE  
CAMBRIDGE, MASSACHUSETTS 02139



Constructive Duality for Discrete Optimization<sup>†</sup>

by

M. L. Fisher<sup>\*</sup> and J. F. Shapiro<sup>\*\*</sup>

WP 625-72

November, 1972

Preliminary Version

The authors would like to acknowledge the helpful comments of M. Held, T. L. Magnanti and M. H. Wagner.

<sup>\*</sup> Graduate School of Business, University of Chicago.

<sup>\*\*</sup> Sloan School of Management, M.I.T.

<sup>†</sup> This research was supported in part by the U.S. Army Research Office (Durham) under Contract No. DA HC04-70-C-0058.

HD28  
N414  
PO 625-12

RECEIVED  
JAN 9 1973  
M. I. T. LIBRARIES

## 1. Introduction

The purpose of this paper is to consolidate and extend recent advances in the use of mathematical programming duality theory in discrete optimization. In particular, meaningful dual programs have been identified for the integer programming problem (Fisher and Shapiro [11]), the resource constrained network scheduling problem (Fisher [10]) and the traveling salesman problem (Held and Karp [22], [23]). The solution techniques developed to exploit the structure of these dual problems are similar and may be applied to a general class of discrete optimization problems. The general problem we consider is: Find

$$\begin{aligned} v(b) = \min f(x) \\ \text{s.t. } g(x) \leq b \\ x \in X \end{aligned} \quad (1.1)$$

where  $X \subseteq R^n$  is a finite set,  $f(x)$  a real valued function and  $g(x)$  a function with values in  $R^m$ . We call problem (1.1) the primal problem.

The dual problem is constructed in the usual fashion (Geoffrion [13]). For  $u \geq 0$ , define the Lagrangean function

$$L(u) = -ub + \min_{x \in X} \{f(x) + ug(x)\}.$$

It is well known and easily shown that  $L(u) \leq v(b)$  and the dual problem is to find the best lower bound; namely, find

$$\begin{aligned} w(b) = \max L(u) \\ \text{s.t. } u \geq 0. \end{aligned} \quad (1.2)$$

By construction,  $w(b) \leq v(b)$  and with the non-convex structure of our primal problem there can be no guarantee that  $w(b) = v(b)$ .

It is often difficult to solve (1.1) directly and some search of the finite set  $X$  is required. Our purpose in constructing and solving one or more dual problems is to make this search efficient by exploiting the following properties of the duals: (1) They are easier to solve than the primal, (2) They



provide lower bounds to the primal problem objective function and those of any primal subproblems derived during the search and (3) they may yield optimal or good feasible solutions to the primal.

The plan of this paper is the following. Section 2 contains four specific applications of the discrete duality theory being studied here. The following section is devoted to a brief review of the properties of dual problems that we use in the construction of algorithms for solving them. Section 3 also contains a brief discussion of the many dual problems in addition to (1.2) which can be constructed from a given primal problem. Section 4 contains a general branch and bound algorithm for solving the primal problem which uses the dual problem (1.2).

The following three sections are devoted to methods for solving a given dual problem. Section 5 contains a primal-dual ascent algorithm for the dual problem. Section 6 contains an approximate search algorithm based on the primal-dual algorithm of section 5 and Scarf's method of simplicial approximation (Scarf [30], Hansen and Scarf [21]). The nature of Scarf's approximation can be given a precise interpretation in this context, and sufficient conditions that the approximation is exact are presented.

The following section contains a brief discussion of subgradient relaxation methods for solving the dual problem which have been studied by Held and Karp [23] and Held, Karp and Wolfe [24].



## 2. Examples of Discrete Dual Problems

In this section we discuss several well-known problems for which the dual problem in Section 1 may be constructed. These problems will illustrate the type of imbedded special structures which give rise to a set  $X$  for which the problem  $\min_{x \in X} f(x) + ug(x)$  is comparatively easy to solve. Minor modifications to the structure of (1.1) are needed for some of these applications, and the corresponding modifications to the dual will be indicated.

Four problems will be discussed: The traveling salesman problem (Held and Karp [22, 23]), the general integer programming (IP) problem (Fisher and Shapiro [11]), the resource-constrained network scheduling problem (Fisher [10]), and the capacitated lot size problem (Dzielinski and Gomory [4], Fisher [9]). The references are to descriptions of dual algorithms developed for these problems. Other problems, too numerous to discuss in detail, to which the dual approach may be applied include the cutting stock problem (Gilmore and Gomory [14]), network synthesis problems (Gomory and Hu [15]) and the problem of scheduling elective hospital admissions (Fisher and Schrage [12]).

### 2.1 The Symmetric Traveling Salesman Problem (Held and Karp [22], [23])

Given a  $k \times k$  matrix of transition costs  $c_{ij}$  with  $c_{ij} = c_{ji}$  we would like to find the least cost tour of  $k$  cities for a traveling salesman. A tour is a path between the cities in which each city is visited once and only once.

The  $k$  cities and the paths between them form an undirected graph with  $k$  nodes and  $\binom{k}{2}$  arcs. Any subgraph of this graph may be represented by the 0-1 variables  $x_{ij}$ ,  $i = 1, \dots, j-1$ ,  $j = 2, \dots, k$ , where  $x_{ij} = 1$  means that the arc  $(ij)$  is present in the subgraph. We let  $x = (x_{12}, x_{13}, x_{23}, \dots, x_{k-1,k})$  denote a vector of these variables.



Following Held and Karp [22] we define a 1-tree as a subgraph formed by adding any two arcs incident on node 1 to a spanning tree for nodes 2, ..., k. A spanning tree for nodes 2, ..., k is a cycless subgraph with at least one arc incident to every node 2, ..., k.

Note that every tour is a 1-tree but that the reverse is not true. If we let  $v_j(x)$  denote the number of arcs incident to node j in the subgraph represented by x, then the tours are exactly the 1-trees which satisfy the side constraints

$$v_j(x) = 2, j = 2, \dots, k - 1$$

The incidence for node 1 is 2 by the definition of a 1-tree, while if nodes 1 to k - 1 have incidence 2, then incidence of 2 for node k is implied.

We may represent the problem of finding a minimum cost tour in the form of problem (1.1) with  $g(x) = b$  by letting

$$X = \{x | \text{the sub-graph defined by } x \text{ is a 1-tree}\},$$

$$f(x) = \sum_{1 \leq i < j \leq k} c_{ij} x_{ij},$$

$$g(x) = (v_2(x), v_3(x), \dots, v_{k-1}(x)),$$

and b is a k-2 vector of 2's.

The problem  $\min_{x \in X} f(x) + ug(x)$  reduces to finding a minimum spanning tree in a graph, a problem relatively easier than the traveling salesman problem. Since the inequalities in (1.1) have been replaced by equalities for this application, the dual variables in (1.2) should be unconstrained in sign.

## 2.2 The Integer Programming (IP) Problem (Gorry and Shapiro [16])

We begin with an IP problem written in the form

$$\min cw$$

$$Aw = d \tag{2.1}$$

$$w \geq 0 \text{ and integer}$$



where  $A$  is an  $n \times (n + m)$  integer matrix of rank  $m$ ,  $d$  is an  $m \times 1$  integer vector, and  $c$  is a  $1 \times (n + m)$  integer vector. Let  $B$  be any  $m \times m$  nonsingular matrix formed from the columns of  $A$  and  $N$  the matrix of remaining columns. We partition  $w$  into  $y$  and  $x$ , with  $y$  corresponding to the columns in  $B$ , and in the same way partition  $c$  into  $c_B$  and  $c_N$ . Then (2.1) may be written as

$$\begin{aligned} \min z &= c_B B^{-1}d + c_N - c_B B^{-1}N)x \\ \text{s.t. } &B^{-1}Nx + y = B^{-1}d \\ &x, y \geq 0 \text{ and integer.} \end{aligned}$$

The correspondence with (1.1) is obtained by letting

$$X = \{x | x \geq 0 \text{ and integer, } y \text{ integer}\},$$

$$f(x) = (c_N - c_B B^{-1}N)x,$$

$$g(x) = B^{-1}N x, \text{ and}$$

$$b = B^{-1}d.$$

The problem  $\min_{x \in X} f(x) + ug(x)$  is equivalent to finding the shortest route in a special network and very efficient algorithms have been developed for the problem (e.g., see Gorry and Shapiro [18]).

It may be shown that if any component of  $c_N - c_B B^{-1}N + uB^{-1}N$  for  $u \geq 0$  is less than zero then  $L(u) = -\infty$ . Hence, it is useful to require  $c_N - c_B B^{-1}N + uB^{-1}N \geq 0$  when we solve the dual. The modifications required to include these linear side constraints in the analysis of the dual problem are straight-forward (Fisher and Shapiro [11]).



### 2.3 The Resource-Constrained Network Scheduling Problem (Fisher [10])

This problem concerns the scheduling of  $I$  jobs to be performed over  $T$  time periods. Each job consists of  $n_i$  tasks and the time to process task  $ij$  is a given integer  $p_{ij}$ . The tasks of each job are constrained by a precedence network defined by a set of tasks  $B_{ij}$  which must be completed before task  $ij$  is started.  $K$  different resources are available, with  $b_{kt}$  denoting the amount of resource  $k$  available in period  $t$ .  $r_{ijk}$  is the amount of resource  $k$  required by task  $ij$  while in process. If  $x_{ij}$  denotes the variable start time of task  $ij$  and  $x = (x_{11}, x_{12}, \dots, x_{In_T})$  we would like to select a value for  $x$  such that some function is optimized. For the sake of concreteness, assume we wish to minimize the sum of the tardiness of each job relative to a due date  $d_i$ .

The correspondence with problem (1.1) is given by

$$X = \{x | 0 \leq x_{ij} \leq T - p_{ij}, x_{ij} \text{ integer, and } x_{ij} \geq x_{i\ell} + p_{i\ell}, \ell \in B_{ij}\},$$

$$f(x) = \sum_{i=1}^I \max(0, \max_j (x_{ij} + p_{ij} - d_i)),$$

$$g_{kt}(x) = \sum_{ij \in J(t)} r_{ijk} \text{ where}$$

$$J(t) = \{ij | x_{ij} + 1 \leq t \leq x_{ij} + p_{ij}\}, \text{ and}$$

$$g(x) = (g_{11}(x), g_{12}(x), \dots, g_{KT}(x)).$$

$X$  is the set of integer start times which satisfy the precedence constraints. The restriction to  $x_{ij}$  integer is allowed because the integrality of  $p_{ij}$  may be used to show that at least one optimal solution will be integer. Letting  $u_{kt}$  denote the multiplier on the constraint  $g_{kt}(x) \leq b_{kt}$ , the problem  $\min_{x \in X} f(x) + ug(x)$  is a CPM problem in which tasks are charged a price  $u_{kt}$  for using resource  $k$  during period  $t$ . An efficient algorithm for this problem is given in [10]. This algorithm uses the separability of the network constraints to facilitate the solution of  $\min_{x \in X} f(x) + ug(x)$ .



## 2.4 The Capacitated Lot Size Problem (Dzielinski and Gomory [4], Fisher [9])

We are given a problem in which  $I$  items are to be produced over  $T$  periods, with a non-negative integer amount  $d_{it}$  demanded of item  $i$  in period  $t$ .  $K$  resources are available in an amount  $b_{kt}$  for resource  $k$  in period  $t$ . We seek to determine integer production amounts  $x_{it}$  for each item in each period. The amount of resource  $k$  used when we produce  $x_{it}$  is given by

$$r_{ik}(x_{it}) = \begin{cases} 0, & x_{it} = 0 \\ a_{ik} + b_{ik}x_{it}, & x_{it} > 0 \end{cases}$$

where  $a_{ik}$  represents setup requirements. Direct per unit production costs for each item are given by  $c_{it}$ . Since demand is fixed, any inventory holding costs may be included in  $c_{it}$ .

The correspondence with (1.1) is:

$$\begin{aligned} x &= (x_{11}, x_{12}, \dots, x_{KT}) \\ X &= \{x | x_{it} \geq 0, \text{ integer, and } \sum_{t=1}^{\tau} x_{it} \geq \sum_{t=1}^{\tau} d_{it} \text{ for all } i \text{ and } \tau\} \\ f(x) &= \sum_i c_{it} x_{it} \\ g_{kt}(x) &= \sum_i r_{ik}(x_{it}) \\ g(x) &= (g_{11}(x), g_{12}(x), \dots, g_{KT}(x)). \end{aligned}$$

The problem  $\min_{x \in X} f(x) + ug(x)$  is separable by item. For any particular item, the problem may be solved using the dynamic programming or shortest route algorithm of Wagner and Whitin [32].



### 3. Properties of Dual Problem

In this section, we review the principal properties of the dual problem which will be used in the remainder of the paper. Let the finite set  $X = \{x^t\}_{t=1}^T$ , and for notational convenience, let

$$\begin{aligned} L(u, x^t) &= -ub + f(x^t) + ug(x^t) \\ &= f(x^t) + u\gamma^t \end{aligned}$$

where  $\gamma^t = g(x^t) - b$ . Then

$$L(u) = \text{minimum}_{t=1, \dots, T} L(u, x^t).$$

It is well known and easily shown that  $L(u)$  is a continuous concave function defined on  $R^m$ . Let

$$Q_t = \{u | u \geq 0 \text{ and } L(u) = L(u, x^t)\};$$

it can easily be shown that  $Q_t$  is a (possibly empty) closed convex set.

With this background, the dual problem (1.2) can be rewritten as the linear programming problem: Find

$$\begin{aligned} w(b) &= \max w \\ \text{s.t. } w &\leq -ub + f(x^t) + ug(x^t) \quad t=1, \dots, T \\ u &\geq 0. \end{aligned} \tag{3.1}$$

This LP formulation of the dual problem immediately gives us some of its properties. The function  $w(b)$  can be shown to be convex and  $w(b) < +\infty$



if and only if  $\{x | g(x) \leq b\} \cap [X] \neq \emptyset$  where  $[X]$  denotes the convex hull of  $X$ .

It will be convenient below to be able to compactify the set of feasible  $u$  being considered in (3.1). The following lemma gives sufficient conditions under which this can be done.

Lemma 3.1. Suppose the point  $x^r \in X$  is an interior point, i.e.  $g_i(x^r) < b_i$ ,  $i=1, \dots, m$ . Then without loss of optimality the constraint  $\sum_{i=1}^m u_i \leq U$  can be added to (3.1) where

$$U = \frac{f(x^r) - w_0}{\min_{i=1, \dots, m} (b_i - g_i(x^r))}$$

and  $w_0$  is a lower bound on  $w(b)$  (e.g.,  $w_0 = L(u)$  for some  $u \geq 0$ ).

Proof: Let  $u^*$ ,  $w^* = w(b)$  by any optimal solution in (3.1). We have

$$\sum_{i=1}^m u_i^* (b_i - g_i(x^r)) \leq f(x^r) - w^* \leq f(x^r) - w_0$$

since  $w^* \geq w_0$ . Division by  $\min_{i=1, \dots, m} (b_i - g_i(x^r)) > 0$  gives us

$$\sum_{i=1}^m u_i^* \leq \sum_{i=1}^m u_i^* \frac{(b_i - g_i(x^r))}{\min_{i=1, \dots, m} (b_i - g_i(x^r))} \leq \frac{f(x^r) - w_0}{\min_{i=1, \dots, m} (b_i - g_i(x^r))}.$$

Thus, the addition of the constraint  $\sum_{i=1}^m u_i \leq U$  does not cut off any optimal solution to (3.1) and the lemma is proven.



Unfortunately, the LP problem (3.1) cannot be efficiently solved as an ordinary LP problem because the number of rows  $T$  is quite large in most applications. Moreover, only a small number of the rows will be binding at an optimal solution. One method of solution of (3.1) is the generalized programming algorithm of Dantzig and Wolfe ([3; Chapter 24]); otherwise known as the dual cutting plane algorithm (Zangwill [34]). In addition to its poor convergence characteristics, this algorithm is not satisfactory for problem (3.1) because it does not provide monotonically increasing lower bounds for the branch and bound algorithm of the next section.

The function  $L(u)$  is not differentiable everywhere but directional derivatives exist in all directions at any point  $u \in R^m$ . Positive directions are the ones to use in ascent algorithms. In order to make these ideas operational, we need some definitions. An  $m$ -vector  $\bar{\gamma}$  is called a subgradient of  $L$  at  $\bar{u}$  if

$$L(u) \leq L(\bar{u}) + (u - \bar{u})\bar{\gamma} \text{ for all } u.$$

Any subgradient of  $L$  at  $\bar{u}$  points into the half space containing all optimal solutions to the dual problem; namely,  $\{u | (u - \bar{u})\bar{\gamma} \geq 0\}$  contains all optimal dual solutions. Thus, it may seem plausible to construct ascent algorithms for the dual problem of the form  $u^{k+1} = u^k + \theta_k \gamma^k$  where  $\gamma^k$  is a subgradient of  $L$  at  $u^L$  and  $\theta_k$  satisfies  $L(u^k + \theta_k \gamma^k) = \max_{\theta \geq 0} L(u^k + \theta \gamma^k)$ . The difficulty is, however, that  $\theta_k$  may be zero but  $u^k$  is not optimal and  $L$  increases in the direction of another subgradient. To put this difficulty in its correct perspective and to see how it can be overcome, we must discuss



briefly the characterization of all subgradients at a point.

Let

$$T(\bar{u}) = \{t \mid L(\bar{u}) = L(\bar{u}, x^t)\};$$

then  $\gamma^t = g(x^t) - b$  for any  $t \in T(\bar{u})$  is a subgradient of  $L$  at  $\bar{u}$  and the collection  $\partial L(\bar{u})$  of all subgradients of  $L$  at  $\bar{u}$  is given by

$$\partial L(\bar{u}) = \{\gamma \mid \gamma = \sum_{t \in T(\bar{u})} \lambda_t \gamma^t, \sum_{t \in T(\bar{u})} \lambda_t = 1, \lambda_t \geq 0, \forall t \in T(\bar{u})\}.$$

With this background, we can give a necessary and sufficient condition in terms of the set  $\partial L(\bar{u})$  that  $\bar{u}$  is optimal in the dual problem. This condition is the basis for our primal-dual ascent algorithm of section five.

Consider a dual solution  $\bar{u} \geq 0$  which we wish to test for optimality in the dual. Let

$$I(\bar{u}) = \{i \mid \bar{u}_i = 0\}$$

and let  $I^c(\bar{u})$  denote the complementary set. The optimality test is derived from the LP problem

$$\begin{aligned} \sigma^* = \min \quad & \sum_{i=1}^m s_i^- + \sum_{i \in I^c(\bar{u})} s_i^+ \\ \text{s.t.} \quad & \sum_{t \in T(\bar{u})} \lambda_t \gamma_i^t - s_i^- + s_i^+ = 0, \quad i=1, \dots, m, \end{aligned} \tag{3.2}$$

$$\sum_{t \in T(\bar{u})} \lambda_t = 1$$

$$\lambda_t \geq 0, \quad t \in T(\bar{u}), \quad s_i^- \geq 0, \quad s_i^+ = 0, \quad i=1, \dots, m.$$



Let  $\lambda_t^*$  denote the optimal weights in (3.2).

Lemma 3.2. The dual solution  $\bar{u} \geq 0$  is optimal in the dual problem (3.1) (equivalently (1.2)) if and only if  $\sigma^* = 0$  in the LP problem (3.2).

Proof: (Sufficiency) If  $\sigma^* = 0$  in (3.2), then the subgradient

$$\sigma^* = \sum_{t \in T(\bar{u})} \lambda_t^* \gamma_t^* \quad (3.3)$$

satisfies  $\gamma_i^* \leq 0$ ,  $i \in I(\bar{u})$ ,  $\gamma_i^* = 0$ ,  $i \in I^c(\bar{u})$ . Thus, for all  $u \geq 0$

$$\begin{aligned} L(u) &\leq L(\bar{u}) + (u - \bar{u})\gamma^* \\ &= L(\bar{u}) + \sum_{i \in I(\bar{u})} (u_i - \bar{u}_i) \gamma_i^* \\ &= L(\bar{u}) + \sum_{i \in I(\bar{u})} u_i \gamma_i^* \\ &\leq L(\bar{u}). \end{aligned}$$

The first inequality follows because  $\gamma^*$  is a subgradient, the first equality because  $\gamma_i^* = 0$ ,  $i \in I^c(\bar{u})$ , the second equality because  $\bar{u}_i = 0$ ,  $i \in I(\bar{u})$  and the second inequality because  $\gamma_i^* \leq 0$ ,  $i \in I(\bar{u})$  and  $u_i \geq 0$ . This establishes sufficiency.

(Necessity) We prove necessity by showing that  $\sigma^* > 0$  in (3.2) implies  $\bar{u}$  is not optimal. To do this, we need to consider the dual LP problem to (3.2)



$$\sigma^* = \max v$$

$$\begin{aligned} \text{s.t. } v &\leq v\gamma^t & t \in T(\bar{u}) \\ -1 &\leq v_i \leq 1 & i \in I^c(\bar{u}) \\ 0 &\leq v_i \leq 1 & i \in I(\bar{u}) \\ v &\text{ unconstrained in sign} \end{aligned} \tag{3.4}$$

Let  $v^*$ ,  $v^* = \sigma^*$  denote an optimal solution to (3.4) and suppose  $v^* > 0$ . This implies that  $v^*\gamma^t > 0$  for all  $t \in T(\bar{u})$  and therefore that  $v^* \neq 0$ . Consider now the problem

$$\max L(\bar{u} + \theta v^*) \tag{3.5}$$

$$\theta \geq 0$$

If the maximum of  $L(u)$  along the half line  $\bar{u} + \theta v^*$  for  $\theta \geq 0$  is greater than  $L(\bar{u})$  or if  $L(u)$  is unbounded along the half line, then clearly  $\bar{u}$  is not optimal in the dual problem and the lemma is proven.

Suppose then that for all  $\theta \geq 0$

$$L(\bar{u} + \theta v^*) \leq L(\bar{u}) \tag{3.6}$$

We will show a contradiction. For  $\theta > 0$ , we have

$$\begin{aligned} L(\bar{u} + \theta v^*) &= \min_{t=1, \dots, T} f(x^t) + \bar{u}\gamma^t + \theta v^*\gamma^t \\ &= \min_{t \notin T(\bar{u})} f(x^t) + \bar{u}\gamma^t + \theta v^*\gamma^t \end{aligned}$$

where the second equality follows from (3.6) and the fact that  $v^*\gamma^t > 0$

$\theta > 0$



for all  $t \in T(\bar{u})$ . Consider a sequence  $\{\theta_k\}_{k=1}^{\infty}$  such that  $\theta_k \rightarrow 0^+$ . Since the number of  $x^t$  is finite, we can assume without loss of generality that there is an  $x^s$ ,  $s \notin T(\bar{u})$ , satisfying for all  $k$

$$L(\bar{u} + \theta_k v^*) = f(x^s) + \bar{u}^s + \theta_k v^{*s} \quad (3.7)$$

Taking limits on both sides of (3.7) and using the continuity of  $L$ , we have

$$L(\bar{u}) = f(x^s) + \bar{u}^s$$

or  $s \in T(\bar{u})$ , a contradiction. This proves necessity and the lemma.

Lemma 3.2 gives necessary and sufficient conditions that a vector  $\bar{u} \geq 0$  is optimal in the dual problem (3.1). It can easily be shown that  $\bar{x} \in X$  satisfying  $g(\bar{x}) \leq b$  is optimal in the primal problem (1.1) if  $L(\bar{u}) = L(\bar{u}, \bar{x})$  and  $\bar{u}(g(\bar{x}) - b) = 0$ . There is no guarantee, however, that such an  $\bar{x}$  exists. Even if one does, it might still be quite difficult to find it since the number of  $x \in X$  satisfying  $L(\bar{u}) = L(\bar{u}, x)$  may be large. In the next section, we show how the dual problem can be used in a "fail-safe" manner as an analytic tool in a general branch and bound algorithm.

Although we have stated only one dual problem (1.2) and its LP equivalent (3.1), there are in fact many possible duals of the primal problem (1.1). A common problem manipulation which can be combined with duality to form many duals is relaxation. A relaxation of (1.1) is a problem of the following form



$$v'(b) = \min f(x)$$

$$\text{s.t. } g_i(x) \leq b_i, \quad i \in I', \quad (3.8)$$

$$x \in X'$$

where  $I' \subseteq \{1, \dots, m\}$ ,  $X \subseteq X'$ . Clearly,  $v(b) \geq v'(b)$  and a dual problem of the form (1.2) with maximum value  $w'(b)$  constructed from (3.8) bears the same relationship to (1.1) as does (1.2) (i.e.  $v(b) \geq v'(b) \geq w'(b)$ ). There is, of course, an increased likelihood that  $v(b) > w'(b)$ , or more generally that  $v(b) - w'(b) > v(b) - w(b) \geq 0$ , but the larger duality gap may be more than compensated by the shorter time it takes to calculate  $w'(b)$ . A more complex form of relaxation, called data relaxation, has been found for integer programming [17]. This type of relaxation consists of a change in the functions  $g_i(x) - b_i$  so that subsequent Lagrangean minimizations can be more efficiently accomplished. We remark that it is possible to use duality theory to try to minimize or reduce the error introduced by the data change (see § 4 of [17]).

Another generalization of the dual problem (1.2) due to Gould [21] is to replace the multiplier vector  $u$  by a multiplier function  $u(\cdot)$  satisfying certain monotonicity properties. The Lagrangean minimization becomes

$$L(u) = \min_{x \in X} \{f(x) + u(g(x) - b)\}.$$

The usefulness of this approach for discrete optimization needs to be carefully studied because it appears difficult to solve the network problems which are the Lagrangean minimizations of section 2 with nonlinear costs. The specific exponential multiplier functions of Evans and Gould [7] are the ones to try in this context.



Finally, we mention briefly that the dual methods here have been widely used to decompose large optimization problems with block diagonal or staircase structures (e.g., Grinold [19], Lasdon [26]). We will not discuss dual decomposition here but simply remark that our methods are applicable to decomposable discrete problems with such structures.



#### 4. Branch and Bound Algorithm for Discrete Primal Problem

As we mentioned in the introduction, the nonconvexity of the set  $X$  makes it possible, and for some problems even likely, that there exist a duality gap. Nevertheless, dual problems can be usefully exploited in a branch and bound algorithm which searches over the finite set  $X$  for an optimal solution to the primal. Our approach bears a resemblance to the branch and bound algorithm of Falk and Soland [8] for a different class of nonconvex programming problems.

The search of the set  $X$  is done in a non-redundant and implicitly exhaustive fashion. At any stage of computation, the least cost known solution  $\hat{x} \in X$  satisfying  $g(\hat{x}) \leq b$  is called the incumbent with incumbent cost  $\hat{z} = f(\hat{x})$ . The branch and bound algorithm generates a series of subproblems of the form

$$\begin{aligned} v(b; X^k) &= \min f(x) \\ \text{s.t. } g(x) &\leq b \\ x &\in X^k \end{aligned} \tag{4.1}$$

where  $X^k \subseteq X$ . If we can find an optimal solution to (4.1), then we have implicitly tested all subproblems of the form (4.1) with  $X^k$  replaced by  $X^\ell \subseteq X^k$  and such subproblems do not have to be explicitly enumerated. The same conclusion is true if we can ascertain that  $v(b; X^k) \geq \hat{z}$  without actually discovering the precise value of  $v(b; X^k)$ . If either of these two cases obtain, then we say that the subproblem (4.1) has been fathomed. If it is not fathomed, then we separate (4.1) into new subproblems of the form (4.1) with  $X^k$  replaced by  $X^\ell$ ,  $\ell=1, \dots, L$ , and

$$\bigcup_{\ell=1}^L X^\ell = X^k, \quad X^{\ell_1} \cap X^{\ell_2} = \emptyset, \quad \ell_1 \neq \ell_2.$$



We attempt to fathom (4.1) by solution of the dual problem

$$\begin{aligned} w(b; X^k) &= \max w \\ \text{s.t. } w &\leq -ub + f(x^t) + ug(x^t) \quad t \in T^k \\ u &\geq 0 \end{aligned} \quad (4.2)$$

where  $T^k \subseteq T$  is the index set of the  $x^t \in X^k$ . The use of (4.2) in analysing (4.1) is illustrated in figure 4.1 which we will now discuss step by step.

STEPS 1 and 2: Often the initial subproblem list consists of only one subproblem corresponding to  $X$ .

STEP 3: A good starting dual solution  $\bar{u} \geq 0$  is usually available from previous computations.

STEP 4: Computing the Lagrangean is a network optimization; shortest route computation for integer programming, minimum spanning tree for the traveling salesman problem, dynamic programming shortest route computation for the capacitated lot size problems, etc.

STEP 5: As a result of step 4, the lower bound  $L(\bar{u}; X^k)$  on  $v(b; X^k)$  is available, where

$$L(\bar{u}; X^k) = -\bar{u}b + \min_{t \in T^k} \{f(x^t) + \bar{u}g(x^t)\} \quad (4.3)$$

It should be clear that (4.1) is fathomed if  $L(\bar{u}; X^k) \geq \hat{z}$  since  $L(\bar{u}; X^k) \leq v(b; X^k)$ .

STEPS 6, 7, 8: Let  $x^s, s \in T^k$ , be an optimal solution in (4.3) and suppose  $x^s$  is feasible; i.e.  $g(x^s) \leq b$ . Since (4.3) was not fathomed (step 5), we have  $L(\bar{u}, X^k) = f(x^s) + \bar{u}(g(x^s) - b) < \hat{z}$ , with the quantity  $\bar{u}(g(x^s) - b) \leq 0$ . Thus, it may or may not be true that  $f(x^s) < \hat{z}$ , but if so, then the incumbent  $\hat{x}$  should be replaced by  $x^s$ . In any case, if  $x^s$  is feasible, we have by duality theory

$$f(x^s) + \bar{u}(g(x^s) - b) \leq v(b; X^k) \leq f(x^s)$$



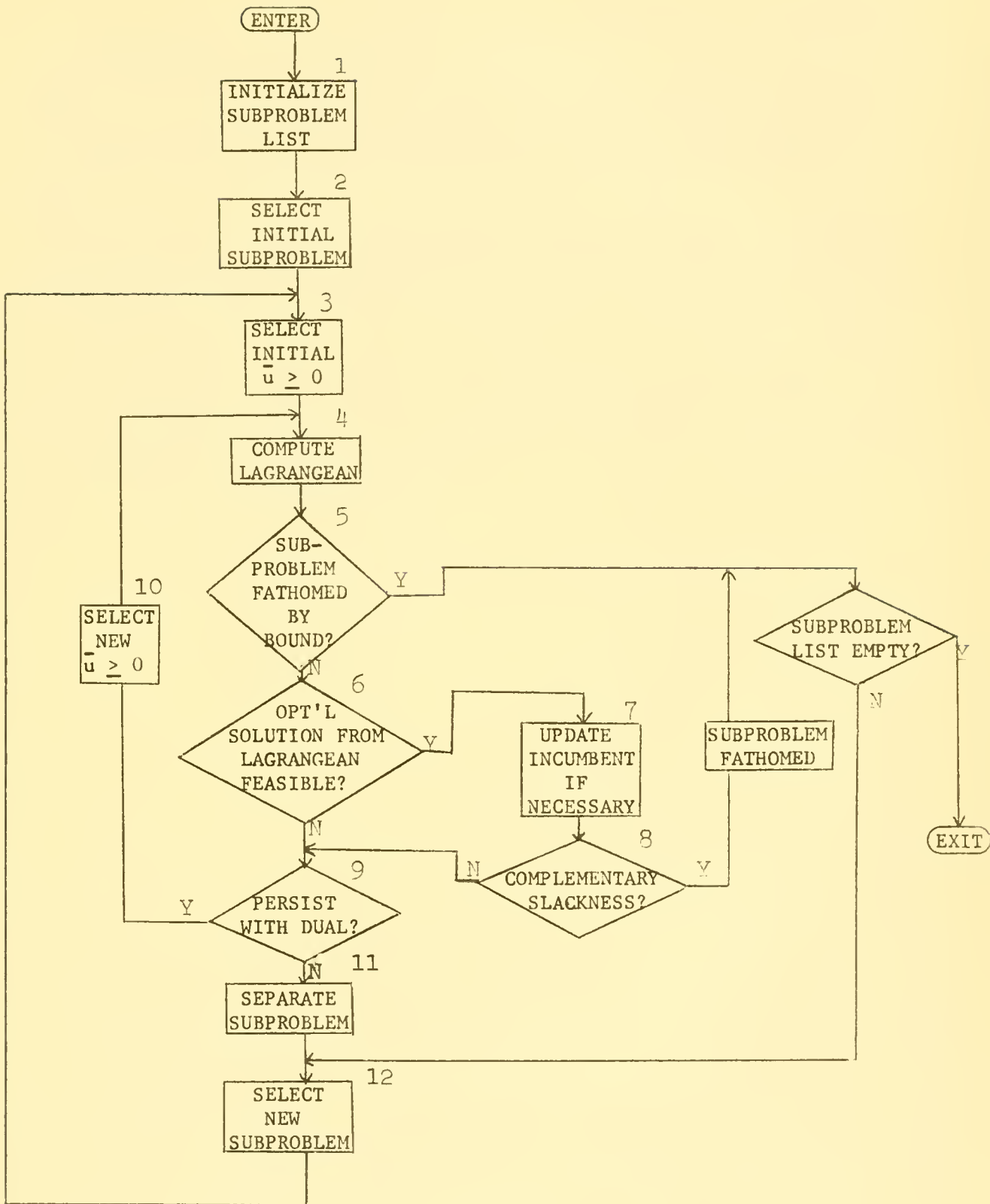


Figure 4.1



and it should be clear that  $x^S$  is optimal if  $\bar{u}(g(x^S)-b) = 0$ ; i.e., if complementary slackness holds.

STEPS 9, 10: These steps are concerned with the functioning of one or more algorithms on the dual problem. For example, step 9 may be a test for optimality on the dual of the current  $\bar{u}$ . Alternatively, it may be a test of recent improvement in the dual lower bound.

STEPS 11, 12: The separation of the given subproblem can often be done on the basis of information provided by the dual problem. For example, in integer programming, problem (4.3) may be separated into two descendants with a zero-one variable  $x_j$  set to zero and one respectively, where  $x_j$  is chosen so that the reduced cost  $f(x_j) + \bar{u}g(x_j)$  is minimal for  $\bar{u}$  optimal. It is important to point out that the greatest lower bound obtained during the dual analysis of (4.3) remains a valid lower bound on a subproblem derived from (4.3) with  $X^l \subseteq X^k$ . In step 12, the new subproblem selected from the subproblem list should be one with low or minimal lower bound.



## 5. Primal-Dual Ascent Algorithm for Dual Problem

In this section we give a primal-dual ascent algorithm for solving the dual problem in the LP form (3.1). The algorithm we propose is an adaptation of the primal-dual simplex algorithm applied to (3.1) where the large number of rows are handled implicitly. Our algorithm bears a close resemblance to the algorithm of Grinold in [20] for large scale linear programming and also to the algorithm of Bradley in [2].

It was demonstrated in lemma 3.2 that the point  $\bar{u} \geq 0$  is optimal in (3.1) if and only if there exists a convex combination  $\gamma^*$  of the subgradients  $\gamma^t$ ,  $t \in T(\bar{u})$ , such that  $\gamma_i^* = 0$  if  $\bar{u}_i > 0$  and  $\gamma_i^* \leq 0$  if  $\bar{u}_i = 0$ . The algorithm proceeds by generating successive  $\gamma^t$ ,  $t \in T(\bar{u})$  until either  $\bar{u}$  is proven optimal in the dual by the test of lemma 3.1 or a new  $\bar{\bar{u}} \neq \bar{u}$  is found such that  $L(\bar{\bar{u}}) > L(\bar{u})$ .

Thus, we consider a generic step of the algorithm beginning at a point  $\bar{u} \geq 0$  and a set  $T'(\bar{u}) \subseteq T(\bar{u})$ . Recall that  $I(\bar{u}) = \{i | \bar{u}_i = 0\}$ . We solve the following LP problem which is simply problem (3.2) with a restricted number of columns:

$$\begin{aligned} \sigma^* &= \min \sum_{i=1}^m s_i^- + \sum_{i \in I^c(\bar{u})} s_i^+ \\ \text{s.t. } \sum_{t \in T'(\bar{u})} \lambda_t \gamma_i^t - s_i^- + s_i^+ &= 0 \quad i=1, \dots, m \\ \sum_{t \in T'(\bar{u})} \lambda_t &= 1 \\ \lambda_t &\geq 0, \quad t \in T'(\bar{u}), \quad s_i^- \geq 0, \quad s_i^+ \geq 0, \quad i=1, \dots, m \end{aligned} \tag{5.1}$$

It is convenient to write the dual of (5.1):



$$\sigma^* = \max v$$

$$\text{s.t. } v \leq v\gamma^t \quad t \in T'(\bar{u}) \quad (5.2a)$$

$$-1 \leq v_i \leq 1 \quad i \in I^C(\bar{u}) \quad (5.2b)$$

$$0 \leq v_i \leq 1 \quad i \in I(\bar{u}) \quad (5.2c)$$

Let  $\sigma^*$  be the optimal solution value in (5.1) and (5.2), and let  $\lambda_t^*$ ,  $t \in T'(\bar{u})$ ,  $v^*$  be optimal values of the primal and dual variables. It should be clear that the direction  $v^*$  produced by (5.2) is somewhat arbitrary; for example, the ranges on the  $v_i$  in (5.2b) and (5.2c) except  $v_i \geq 0$  in (5.2c) could be adjusted to reflect differences in scaling on different rows. The reader is referred to Nemhauser and Widhelm [28] for more discussion of this point.

By lemma 3.2,  $\bar{u}$  is optimal in the dual if  $\sigma^* = 0$ . We describe a generic step of the algorithm for the case  $\sigma^* > 0$ . Notice from (5.2) that for at least one  $t \in T'(\bar{u})$ , we must have  $v^*\gamma^t = \sigma^* > 0$ . Thus, we have found a subgradient  $\gamma^* = \sum_{t \in T'(\bar{u})} \lambda_t^* \gamma^t$  and a direction  $v^* \neq 0$  such that an increase in the Lagrangean function is indicated in the direction  $v^*$  namely, for  $\theta \geq 0$ ,

$$L(\bar{u} + \theta v^*) \leq L(\bar{u}) + \theta v^* \gamma^* = L(\bar{u}) + \theta \sigma^* \quad (5.3)$$

The primal-dual algorithm proceeds by finding the maximal value of  $\theta$  such that equality holds in (5.3).

Let

$$V(\bar{u}) = \{i | v_i^* < 0\};$$



note that  $V(\bar{u}) \subseteq I^c(\bar{u})$ . The search for a maximal  $\theta$  with the desired property is confined to the interval  $[0, \theta_{\max}]$  where

$$\theta_{\max} = \min \left\{ \frac{\bar{u}_i}{-v_i^*} : i \in V(\bar{u}) \right\} > 0.$$

This interval is the one in which  $u$  remains non-negative. It is possible, of course, that  $V(\bar{u}) = \emptyset$ , and no upper bound on  $\theta$  is available. As long as  $w(b) < +\infty$ , this does not cause any real difficulty because we can easily devise a finite search of the half line  $\bar{u} + \theta v^*$ , for  $\theta \geq 0$ , for the maximum of  $L$  along it. Alternatively, we could add a

constraint  $\sum_{i=1}^m u_i \leq U$  to the dual problem. A valid  $U$  can be derived if

an interior point  $x^r \in X$  exists (see lemma 3.1).

In any

case, an arbitrary upper bound on  $U$  can be selected. The only inviolate use we make of the dual problem in the algorithm of section 4 for solving the given primal problem is to provide lower bounds. The addition of

a constraint  $\sum_{i=1}^n u_i \leq U$  to the dual (3.1) does not affect this property.

Henceforth, we shall assume that the above difficulty is handled by the imposition of the constraint  $\sum_{i=1}^m u_i \leq U$ . Let  $\theta^* \in [0, \theta_{\max}]$  denote the maximal value of  $\theta$  such that  $L(\bar{u} + \theta v^*) = L(\bar{u}) + \theta \sigma^*$ . The procedure for finding the maximum is the following. Let  $\theta_1 = \theta_{\max}$  and at iteration  $k$ , compute  $L(\bar{u} + \theta_k v^*) = L(\bar{u}, x^k)$  for some  $x^k \in X$ . Notice that  $L(\bar{u} + \theta_k v^*) = L(\bar{u}, x^k) + \theta_k v^* \gamma^k$ . If  $L(\bar{u} + \theta_k v^*) = L(\bar{u}) + \theta_k v^* \gamma^*$ , then  $\theta_k = \theta^*$  and the search terminates. If

$$L(\bar{u} + \theta_k v^*) < L(\bar{u}) + \theta_k v^* \gamma^*, \quad (5.4)$$



then compute

$$\theta_{k+1} = \frac{L(u, x^k) - L(\bar{u})}{v^* \gamma^* - v^* \gamma^k} \quad (5.5)$$

The denominator in this fraction is positive because of (5.4) and the fact that  $L(\bar{u}, x^k) \geq L(\bar{u})$ . If  $\theta_{k+1} = 0$ , then set  $\theta^* = 0$  and the search terminates.

We show now that (i) the solution of the LP (5.1) defined at  $\bar{u} + \theta^* v^*$  can begin with the optimal basis for (5.1) at  $\bar{u}$  and (ii) the value of the objective function at  $u + \theta^* v^*$  will be (barring degeneracy) less than  $\sigma^*$ . Convergence of the primal-dual algorithm is thereby assured by the usual simplex criterion (see Simonnard [32; pp. 128-134]). There are three cases to consider.

Case one:  $0 < \theta^* < \theta_{\max}$

In this case, we want to show that the optimal basic variables from (5.1) at  $\bar{u}$  remain in the problem at  $\bar{u} + \theta^* v^*$  and with the same objective function coefficients. First, for  $t \in T'(\bar{u})$  such that  $\lambda_t$  is in the optimal basis for (5.1), we have by complementary slackness (see (5.2)) that  $v^* \gamma^t = \sigma^*$ . This implies that  $t \in T(\bar{u} + \theta^* v^*)$  since by construction,  $L(\bar{u} + \theta^* v^*) = L(\bar{u} + \theta^* v^*, x^t) = L(\bar{u}) + \theta^* v^* \gamma^t = L(\bar{u}) + \theta^* \sigma^*$ .

To complete the argument, we need to consider the  $s_i^-$  and  $s_i^+$  variables. They clearly remain in (5.1) at  $\bar{u} + \theta^* v^*$  and can be in a starting feasible basis. We must simply show that the objective is unchanged for the  $s_i^+$  that are basic. For  $i \in I^c(\bar{u})$ , we have that  $I^c(\bar{u}) \subseteq I^c(\bar{u} + \theta^* v^*)$  because  $\theta^* < \theta_{\max}$ ; the coefficient of  $s_i^+$  for such an  $i$  thus remains equal to 1. Consider  $s_i^+$  for  $i \in I(\bar{u})$ . If  $s_i^+$  is basic in (5.1), then by complementary slackness,  $v_i^* = 0$  which implies  $i \in I(\bar{u} + \theta^* v^*)$  and the objective function



coefficient remains equal to zero. In short, we use the previously optimal basis as a starting feasible basis for (5.1) at  $\bar{u} + \theta^* v^*$  and the objective function value equals  $\sigma^*$ .

To show that this value of  $\sigma^*$  can be reduced for (5.1) at  $\bar{u} + \theta^* v^*$ , note that  $\theta^* > 0$  means we have found an  $x^s \in X$  such that  $L(\bar{u}, x^s) > L(\bar{u}) = L(\bar{u}, x^t)$  for  $\lambda_t$  basic in the starting basis. By construction,  $L(\bar{u} + \theta^* v^*) = L(\bar{u}, x^s) + \theta^* v^* \gamma^s = L(\bar{u}) + \theta^* \sigma^*$  and therefore  $v^* \gamma^s < \sigma^*$ . In other words, the LP reduced cost relative to the starting basis of the subgradient  $\gamma^s$  of  $L$  at  $\bar{u} + \theta^* v^*$  is  $v^* \gamma^s - \sigma^* < 0$ . The activity  $\gamma^s$  should be pivoted into the starting basis and (barring degeneracy) the objective function will be decreased.

Case two:  $\theta^* = \theta_{\max}$

This case is the same as case one except for the possibility that for some  $i \in I^c(\bar{u})$ , we have  $s_i^+$  basic and  $\bar{u}_i + \theta^* v_i = 0$  implying  $i \in I(\bar{u} + \theta^* v)$ . If this is so, the objective function coefficient of  $s_i^+$  needs to be changed from 1 to 0 which causes a reduction in the objective function. Problem (5.1) is then reoptimized.

Case three:  $\theta^* = 0$

In this case, we have found an  $s \in T(\bar{u})$  such that  $L(\bar{u}, x^s) = L(\bar{u})$  and  $v^* \gamma^* - v^* \gamma^s > 0$  (see (5.5)). Thus,  $v^* \gamma^s < \sigma^*$ , we have  $s \in T(\bar{u}) - T'(\bar{u})$  and the non-basic  $\gamma^s$  can enter the basis causing the objective function to be reduced.

The primal-dual rule of moving to the next breakpoint of the  $L$  function in the direction  $v^*$  ensures convergence of the ascent algorithm. It is also reasonable to move to the maximum of  $L$  in the direction  $v^*$ , but then convergence cannot be ensured. This is a curious point which needs more study from both a theoretical and computational viewpoint (see also Grinold [19; p. 452]).



## 6. Simplicial Approximation Algorithm for Dual Problem

In this section, we present a search method for approximating an optimal solution to the dual problem (1.2) by approximating the optimality conditions of lemma 3.2. The search uses Scarf's method for simplicial approximation of fixed points [30]. Since Scarf's original work in reference 30 on fixed point approximation, there have been a number of important theoretical and computational extensions (Hansen and Scarf [21], Eaves [5,6], Kuhn [25]). For expositional reasons, however, we will ignore these extensions here and merely indicate where they are relevant and necessary for efficient computation.

Our method draws on some of the construction of Hansen and Scarf for simplicial approximation in non-linear programming [21; pp. 14-18]. The major conceptual difference between our method and theirs is that we are approximating optimal dual solutions while they are approximating optimal primal solutions. Wagner [33] has also applied simplicial approximation to nonlinear programming.

In order to compactify the search, we must append a constraint of the form  $\sum_{i=1}^m u_i \leq U$  to the dual problem which we will consider in the LP form (cf (3.1))

$$w(b) = \max w$$

$$w \leq -ub + f(x^t) + ug(x^t) \quad t=1, \dots, T$$

$$\sum_{i=1}^m u_i \leq U \tag{6.1}$$

$$u \geq 0.$$



We have already discussed in lemma 3.1 the validity of adding such a constraint to the dual problem. The necessary and sufficient condition for optimality of  $\bar{u}$  in (6.1) from lemma 3.1 becomes  $\sigma^* = 0$  in

$$\begin{aligned} \sigma^* &= \min \sum_{i=1}^m s_i + \sum_{i \in I^+(\bar{u})} s_i^+ \\ \text{s.t. } \sum_{t \in T(\bar{u})} \lambda_t + \gamma_i^+ - s_i^- + s_i^+ - \theta &= 0 \end{aligned} \quad (6.2)$$

$$\sum_{t \in T(\bar{u})} \lambda_t = 1$$

$$\lambda_t \geq 0, t \in T(\bar{u}), s_i^- \geq 0, s_i^+ \geq 0, i=1, \dots, m$$

where

$$\begin{aligned} \theta &= 0 & \text{if } \sum_{i=1}^m \bar{u}_i < U \\ \theta &\geq 0 & \text{if } \sum_{i=1}^m \bar{u}_i = U. \end{aligned}$$

For technical reasons, it is convenient to add a zero<sup>th</sup> component to the  $u$  vector and consider our search to be over the simplex

$$S = \{u \mid \sum_{i=0}^m u_i = U\}. \quad (6.3)$$

This modification holds only for this section. In [30], Scarf picks points in the set  $P = \{u^0, u^1, \dots, u^K\}$  where  $u^k \in S$ ,  $k > m$  and where



$$u^0 = (0, M_0, \dots, M_0)$$

$$u^1 = (M_1, 0, M_1, \dots, M_1)$$

$$\vdots$$

$$u^m = (M_m, \dots, M_m, 0)$$

and  $M_0 > M_1 > \dots > M_m > U$ . In addition, the  $u^k$  are chosen so that

$u_i^{k_1} \neq u_i^{k_2}$  for any  $i$  and any  $k_1 \neq k_2$ ,  $k_1 > m$  and  $k_2 > m$ . With the points  $u^k$  so chosen, Scarf defines [30; p. 1330] a primitive set  $u^{k_0}, \dots, u^{k_m}$  to

be one such that there is no  $u^k$  satisfying  $u_i^k \geq \min_{i=0,1,\dots,m} (u_i^{k_i})$ ,

$i=0, \dots, m$ . In words, a primitive set describes a subsimplex in which no other  $u^k$  lie. From a computational viewpoint, the simplex  $S$  can be partitioned into a regular grid of subsimplices and moreover, the points  $u^k$  can be generated as required (see Hansen and Scarf [21], Eaves [5]).

The main tool in Scarf's simplicial approximation algorithm is the following result.

Lemma 6.1 (Scarf [30; p. 1332]) Let  $u^{k_0}, \dots, u^{k_m}$  be a primitive set, and let  $u^{k_\alpha}$  be a specific one of these vectors. Then, aside from one exceptional case, there is a unique vector  $u^k$  different from  $u^{k_\alpha}$  and such that  $u^{k_0}, \dots, u^{k_{\alpha-1}}, u^k, u^{k_{\alpha+1}}, \dots, u^{k_m}$ , form a primitive set. The exceptional case occurs when the  $m$  vectors  $u^{k_i}$ ,  $i \neq \alpha$ , are all selected from  $u^0, \dots, u^m$ , and in this case no replacement is possible.

To apply this result to solving approximately (6.1), we need to apply another theorem which extends the above result.

Theorem 6.1 (Hansen and Scarf [21; p. 3]) Let



$$A = \begin{pmatrix} 1 & 0 & 0 & a_{0,m+1} & a_{0,K} \\ 0 & 1 & \cdot & \cdot & \cdot \\ \cdot & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & \cdot & \cdot \\ 0 & 0 & 1 & a_{m,m+1} & a_{m,K} \end{pmatrix}$$

be an  $(m+1) \times (K+1)$  matrix and  $b^T = (b_0, \dots, b_m)$  a  $(m+1)$  non-negative vector, such that the  $k^{th}$  column of  $A$  is associated with the  $k^{th}$  vector  $u^k \in P$ . Assume that the set of non-negative vectors  $y$ , satisfying  $Ay = b$ , is bounded. Then there exists a primitive set  $u^{k_0}, \dots, u^{k_m}$  so that the columns  $k_0, \dots, k_m$  form a feasible basis for  $Ay = b$ .

The theorem is applied to (6.1) by specifying the appropriate matrix  $A$  and vector  $b$ . The matrix  $A$  is given below and the appropriate  $u^k$  are shown above their associated columns.

$$A = \begin{matrix} & \begin{matrix} u^0 & u^1 & & u^m & u^{m+1} & & u^K \end{matrix} \\ \begin{pmatrix} 1 & 0 & & 0 & 1 & & 1 \\ 0 & 1 & & \cdot & \gamma_1^{m+1} + 1 & & \gamma_1^K + 1 \\ \cdot & 0 & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & 0 & \cdot & & \cdot \\ 0 & 0 & & 1 & \gamma_m^{m+1} + 1 & & \gamma_m^K + 1 \end{pmatrix} \end{matrix}$$

where  $\gamma^k$  is any subgradient of  $L$  at  $u^k$ . The vector  $b^T = (1, \dots, 1)$ .

Clearly, the set  $\{y | Ay=b, y \geq 0\}$  is bounded in this case.



We sketch the constructive proof of theorem 6.1 because it implies an algorithm which we need to discuss in some detail. After the discussion of the algorithm we will give a discussion of the results of the algorithm and their interpretation as an approximation to the optimality conditions of lemma 3.1.

The algorithm begins with the set  $u^0, u^1, \dots, u^m$  which is not a primitive set. A primitive set is uniquely formed by replacing  $u^0$  with the vector  $u^{k*}$  where

$$u_0^{k*} = \max_{k > m} u_0^k.$$

We then take the feasible basis corresponding to the first  $m+1$  columns of  $A$  and pivot in column  $k^*$ . This pivot is the ordinary simplex pivot. If the column of  $A$  corresponding to  $u^0$  is pivoted out of the starting basis then the algorithm terminates. If another column is pivoted out of the starting basis, say column  $k_1$ , then the primitive set  $u^{k*}, u^1, \dots, u^m$  and the new basis correspond except for  $u^1$  and the zero<sup>th</sup> column of  $A$  which is still in the basis. In this case, we remove  $u^1$  from the primitive set, and replace it with a unique  $u \in P$  according to lemma 6.1. If  $u^0$  replaces it, then the algorithm terminates. Otherwise, the algorithm continues by repeating these same operations.

At an arbitrary iteration, we have a primitive set  $u^{k_0}, u^{k_1}, \dots, u^{k_m}$  and a feasible basis corresponding to the columns  $0, k_1, \dots, k_m$ .

(i) Pivot column  $k_0$  into the basis. If column 0 is pivoted out of the basis, terminate the algorithm. Otherwise, go to (ii).



(ii) If column  $k_t$  was pivoted out of the basis in step (i), take  $u^{k_t}$  out of the primitive set. If  $u^0$  comes into the primitive set, terminate the algorithm. Otherwise, repeat (i).

The finiteness of the algorithm is due to the uniqueness of the changes at each of the above steps and the finiteness of the set  $P$  and the possible bases of  $Ay = b$ .

Thus, the algorithm implicit in theorem 6.1 terminates with a primitive set  $u^{k_0}, \dots, u^{k_m}$  and a non-negative basic solution of  $Ay = b$  using the basis columns  $k_0, \dots, k_m$  of  $A$ . Let  $s_0, \dots, s_m$  denote the variables corresponding to the first  $m+1$  columns of  $A$ , and let  $\lambda_{m+1}, \dots, \lambda_K$  denote the remaining variables. Let  $\tilde{s}_i$  and  $\tilde{\lambda}_k$  denote the specific values of these variables in the terminal solution; we want to demonstrate that the terminal solution approximates the necessary and sufficient condition for optimality that  $\sigma^* = 0$  in (6.2). The condition is approximated because the subgradients in the desired convex combination are subgradients of  $L$  chosen not at a single point but from  $m+1$  points which are close together. There are two cases to consider.

Case one:  $k_0 \neq 0$  (algorithm terminated at step (i)):

In this case, we have  $\tilde{s}_0 = 0$  and row 0 of  $Ay = b$  gives us

$$\sum_{k=m+1}^K \tilde{\lambda}_k = 1 \quad (6.4)$$

Recall that  $\tilde{\lambda}_k > 0$  for those  $u^k$  which are close together in the final primitive set. For row  $i$ ,  $i=1, \dots, m$ , we have



$$\tilde{s}_i + \sum_{k=m+1}^K (\gamma_i^k + 1) \tilde{\lambda}_k = 1,$$

and using (6.4), we have

$$\tilde{s}_i + \sum_{k=m+1}^K \gamma_i^k \tilde{\lambda}_k = 0. \quad (6.5)$$

If  $\tilde{s}_i > 0$ , then  $u^i$  is in the final primitive set and  $u_i^j \approx 0$ ,  $j=0, \dots, m$  and  $\sum_{k=m+1}^K \gamma_i^k \tilde{\lambda}_k = -\tilde{s}_i < 0$  is the appropriate approximate optimality condition for row  $i$ . On the other hand, if  $u^i$  is not in the final primitive set,  $s_i$  is not a basic variable implying  $\tilde{s}_i = 0$  and  $\sum_{k=m+1}^K \gamma_i^k \tilde{\lambda}_k = 0$  which is again the appropriate approximate optimality condition.

Case two:  $k_0 = 0$  (algorithm terminated at step (ii))

In this case, the algorithm has progressed from the region of  $S$  where  $\sum_{i=1}^m u_i \approx 0$  to the boundary  $\sum_{i=1}^m u_i \approx U$ . By the definition of primitive sets, there must be some  $u^\ell$ ,  $1 \leq \ell \leq m$ , not in the final primitive set implying  $\tilde{s}_\ell = 0$ . Since the right hand side on row  $\ell$  is 1, there must be at least one  $\tilde{\lambda}_k > 0$  and therefore  $\sum_{k=m+1}^K \tilde{\lambda}_k > 0$ . We normalize the non-negative weights to sum to one by taking

$$\hat{\lambda}_k = \frac{\tilde{\lambda}_k}{\sum_{k=m+1}^K \tilde{\lambda}_k}, \quad k=m+1, \dots, K.$$

From row 0, we have  $\sum_{k=m+1}^K \tilde{\lambda}_k = 1 - \tilde{s}_0 \leq 0$ , where the inequality is strict (barring degeneracy) because  $\tilde{s}_0$  is basic. For row  $i$ , we have



$$\tilde{s}_i + \sum_{k=m+1}^K \gamma_i^k \tilde{\lambda}_k = 1 - \sum_{k=m+1}^K \tilde{\lambda}_k = \tilde{s}_0 \geq 0,$$

and dividing by  $\sum_{k=m+1}^K \tilde{\lambda}_k$ ,

$$\frac{1}{\sum_{k=m+1}^K \tilde{\lambda}_k} \tilde{s}_i + \sum_{k=m+1}^K \gamma_i^k \hat{\lambda}_k - \frac{\tilde{s}_0}{\sum_{k=m+1}^K \tilde{\lambda}_k} = 0 \quad (6.6)$$

Since the  $u_i^j$ ,  $j=0,1,\dots,m$ , in the final primitive set all satisfy  $\sum_{i=1}^m u_i^j \approx U$ , it is appropriate to allow the non-negative quantity

$$\tilde{\theta} = \frac{\tilde{s}_0}{\sum_{k=m+1}^K \tilde{\lambda}_k}$$

on every row  $i$ ,  $i=1,\dots,m$ . As before,  $\tilde{s}_i > 0$  implies  $u^i$  in the final primitive set and the condition  $\sum_{k=m+1}^K \gamma_i^k \hat{\lambda}_k - \tilde{\theta} \leq 0$  is appropriate.

The above analysis establishes the nature of the approximation in terms of approximating optimality conditions. It is also natural to ask: what is an approximately optimal solution  $\tilde{u}$  in (6.1)? Since  $u^{k_0}, \dots, u^{k_m}$  are close together, it is reasonable to take any one of them or any convex combination of them. The best choice, however, appears to be any optimal solution in the linear programming problem



$$\begin{aligned}
 \tilde{w}(b) &= \max w \\
 \text{s.t. } w &\leq -ub + f(x^t) + ug(x^t), \quad t \in \tilde{T} \\
 \sum_{i=1}^m u_i &\leq U \\
 u &\geq 0
 \end{aligned} \tag{6.7}$$

where  $\tilde{T}$  is the index set of the  $\gamma^t$  corresponding to basic  $\lambda$  variables in the terminal basic feasible solution. We have argued above that  $\tilde{T}$  contains at least one  $t$ .

Let  $\tilde{u}$  be an optimal solution in (6.7) and compute  $L(\tilde{u})$ . It is easy to see that  $\tilde{u}$  is optimal in the dual problem (6.1) if  $\tilde{w}(b) = L(\tilde{u})$ . This is because  $L(\tilde{u}) \leq -\tilde{u}b + f(x^t) + \tilde{u}g(x^t)$  for all  $t$  and therefore the constraints from (6.1) omitted in forming (6.7) are satisfied a fortiori. In any case,  $\tilde{w}(b) \geq w(b)$  and  $\tilde{w}(b) - L(\tilde{u}) \geq 0$  is an upper bound on the objective function error of stopping with  $\tilde{u}$ .

Let us now consider briefly the convergence properties of the simplicial approximation as the  $u^k$ ,  $k > m$  are increased in number and the diameter of the subsimplices defined by primitive sets approach zero. For  $r = 1, 2, \dots$ , let  $u^{k_0, r}, \dots, u^{k_m, r}$  denote the defining points for the terminal primitive set at the  $r^{\text{th}}$  partition of  $S$ . As  $r$  goes to infinity, the diameter of these subsimplices goes to zero, and there is a subsequence of the  $u^{k_0, r}, \dots, u^{k_m, r}$  converging to a single point  $u^*$ . Any such point must be optimal in (6.1). To see this, recall that there is only a finite number of distinct columns possible for the system  $Ay = b$ . Thus, there is some basic set repeated infinitely often in the converging subsequence. Since the sets  $Q_t$  defined in section 3 are closed, the approximate optimality conditions represented by this fixed basic set become exact at  $u^*$ .



An extremely important extension which we have ignored is the procedure for Eaves [6] for restarting the simplicial approximation search with a finer grid using the terminal solution obtained with the current grid. A second construction which may be very important if simplicial approximation is to be computationally feasible is the variation of Kuhn in [25] which allows the dimension of the search at each iteration to be limited to the number of positive multipliers. Simplicial approximation has been computationally attractive on problems of up to 40 dimensions. It is conceivable that a simplicial approximation search of a 300 row integer programming problem in the form of section 2.2 would involve no more than 40 positive multipliers (i.e. 40 binding constraints) at any iteration.



## 7. Subgradient Relaxation Methods for Dual Problem

In this section we discuss the very interesting work of Held and Karp [23] and Held and Karp and Wolfe [24] on the application and extension of "the relaxation method for linear inequalities" of Agmon [1] and Motzkin and Schoenberg [27] to the traveling salesman and other optimization problems. We will refer to this method as subgradient relaxation to distinguish it from the data and problem relaxation methods discussed in section 3. Our discussion in this section will be brief and the development will follow the development of [23] incorporating an observation from [24].

The idea behind subgradient relaxation is to generate a sequence  $\{u^r\}$  of non-negative dual variables by the rule

$$u_i^{r+1} = \max\{u_i^r + \theta_r \gamma_i^r, 0\} \quad i=1, \dots, m,$$

where  $\gamma^r$  is any subgradient of  $L$  at  $u^r$  and  $\theta_r$  is a positive constant to be chosen from a specified range. There is no guarantee that  $L(u^{r+1}) > L(u^r)$  but if  $w(b)$  is known, the method can be made to converge (in a certain case, finitely) to a  $\bar{u}$  such that  $L(\bar{u}) \geq \bar{w}$  where  $\bar{w}$  is arbitrarily close to  $w(b)$ . Of course,  $w(b)$  is generally not known although it can often be closely estimated. Another point worth noting is that in using the subgradient relaxation method in the branch and bound algorithm of figure 4.1, we can take  $\bar{w}$  to be the lower bound required for fathoming the subproblem by bound in step 5. In any case, the justification for the subgradient relaxation method rests with its computation effectiveness. The method has proven effective for the traveling salesman problem [23].



Consider now a generic step from the point  $u \geq 0$  to the point  $u + \theta\gamma + \delta$  where  $\theta$  is some positive number to be chosen,  $\gamma$  is a subgradient of  $L$  at  $u$  and  $\delta_i = 0$  if  $u_i + \theta\gamma_i \geq 0$ ,  $\delta_i = -u_i - \theta\gamma_i < 0$  if  $u_i + \theta\gamma_i < 0$ ,  $i=1, \dots, m$ . The following lemma stated without proof summarizes the results on page 9 of [23] and page 509 of [24]. In words, it says that if a sufficiently small step in the direction of the subgradient  $\gamma$  is taken, followed by a projection onto the non-negative orthant, then the new point  $u + \theta\gamma + \delta$  is closer to a maximal  $u^*$  than  $u$ .

Lemma 7.1 (Held, Karp and Wolfe): Let  $u \geq 0$  with dual value  $L(u)$  be given. Suppose  $\bar{u}$  is such that  $\bar{u} \geq 0$  and  $L(\bar{u}) > L(u)$ . If  $\theta$  is chosen to satisfy

$$0 < \theta < \frac{2(L(\bar{u}) - L(u))}{\|\gamma\|^2},$$

then

$$\|\bar{u} - (u + \theta\gamma + \delta)\| < \|\bar{u} - u\|.$$

This lemma indicates the type of step size required for convergence. Assume  $\bar{w} < w(b)$  and let  $U_{\bar{w}}^m$  denote the set of  $u \in R^m$  satisfying

$$\begin{aligned} \bar{w} &\leq -ub + f(x^t) + ug(x^t) & t=1, \dots, T \\ u &\geq 0. \end{aligned} \tag{7.1}$$

Let  $\{u^r\}$  be a sequence of points given by

$$u_i^{r+1} = \max\{u_i^r + \rho_r \left( \frac{\bar{w} - L(u^r)}{\|\gamma^r\|^2} \right) \gamma_i, 0\} \tag{7.2}$$



The results of Agmon [1] and Motzkin and Schoenberg [27] are concerned with convergence properties of this sequence. Roughly speaking, if  $\rho_r = 2$  for all  $r$ , the sequence always includes a point  $u^{\ell} \in U_{\bar{w}}$  whereas if  $0 < \epsilon < \rho_r \leq 2$  for all  $r$ , the sequence includes a point  $u^{\ell} \in U_{\bar{w}}$  or converges to a point on the boundary of  $U_{\bar{w}}$ .

The advantages of subgradient relaxation over the primal-dual ascent algorithm of section 5 is the elimination of computational overhead while maintaining satisfactory or good convergence properties. One disadvantage is the absence of guaranteed monotonically increasing lower bounds. Moreover, when there are constraints other than  $u \geq 0$  on the dual problem (1.2) (e.g., the IP dual in [11]), the subgradient relaxation method requires that the point  $u + \theta\gamma$  be projected in a non-trivial fashion onto the set of dual feasible solutions. Computational experiments are underway to compare and integrate the primal-dual ascent algorithm and subgradient relaxation method for a variety of discrete optimization problems.



### References

1. S. Agmon, "The relaxation method for linear inequalities," Canadian Journal of Mathematics 6 (1954), 382-392.
2. S. P. Bradley, "Decomposition programming and economic planning," ORC Rep. 67-20, Operations Research Center, University of California, Berkeley, 1967.
3. G. B. Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton, N.J., 1963.
4. B. P. Dzielinski, and R. Gomory, "Optimal Programming of Lot Size Inventory and Labor Allocations," Management Science 11 (1965), 874-890.
5. B. Curtis Eaves, "Computing Kakutani Fixed Points," SIAM J. of Appl. Math. 21 (1971), 236-244.
6. B. C. Eaves, "Homotopies for Computation of Fixed Points," Mathematical Programming 3 (1972), 1-20.
7. J. P. Evans and F. J. Gould, "Stability and Exponential Penalty Function Techniques in Nonlinear Programming," Institute of Statistics Mimeo Series No. 723, U. of North Carolina, (1970).
8. J. E. Falk and R. M. Soland, "An Algorithm for Separable Nonconvex Programming Problems," Management Science (1969), 550-569.
9. M. L. Fisher, "Optimal Solution of Resource Constrained Network Scheduling Problems," Technical Report No. 56, Operations Research Center, Massachusetts Institute of Technology, Cambridge, Mass., September, 1970.
10. M. L. Fisher, "Optimal Solution of Scheduling Problems Using Generalized Lagrange Multipliers and Relaxation," Report 720, Center for Mathematical Studies in Business and Economics, University of Chicago, February, 1971. (To appear in Operations Research.)
11. M. L. Fisher and J. F. Shapiro, "Constructive Duality in Integer Programming," working paper OR 008-72, Operations Research Center, Massachusetts Institute of Technology, Cambridge, Mass., April, 1972.
12. M. L. Fisher and Linus Schrage, "Using Lagrange Multipliers to Schedule Elective Hospital Admissions," University of Chicago Report, July, 1972.
13. A. M. Geoffrion, "Duality in Nonlinear Programming: A Simplified Applications-Oriented Development," SIAM Review 13 (1971), 1-37.
14. P. C. Gilmore and R. E. Gomory, "A Linear Programming Approach to the Cutting-Stock Problem, Part II," Operations Research, 11 (1963), 863-888.



15. R. E. Gomory and T. C. Hu, "Synthesis of a Communication Network," SIAM J. 12 (1964), 348-389.
16. G. A. Gorry and J. Shapiro, "An Adaptive Group Theoretic Algorithm for Integer Programming Problems," Management Science 17 (1971), 285-306.
17. G. A. Gorry and J. F. Shapiro and L. A. Wolsey, "Relaxation Methods for Pure and Mixed Integer Programming Problems," Management Science 18 (Jan., Part I, 1972), 229-239.
18. G. A. Gorry and J. F. Shapiro, "Computational Experience with a Group Theoretic Integer Programming Algorithm," WP 603-72, Sloan School of Management, May 1972 (to appear in Mathematical Programming).
19. Richard C. Grinold, "Steepest Ascent for Large Scale Linear Programs," SIAM Review 14 (1972), 447-464.
20. F. J. Gould, "Extensions of Lagrange Multipliers in Nonlinear Programming," SIAM Journal on Applied Mathematics 17 (1969), 1280-1297.
21. T. Hansen and H. Scarf, "On the Applications of a Recent Combinatorial Algorithm," Cowles Foundation Discussion Paper No. 272, Yale University, New Haven, Connecticut, (1969).
22. M. Held and R. M. Karp, "The traveling salesman problem and minimum spanning trees," Operations Research 18 (1970), 1138-1162.
23. M. Held and R. M. Karp, "The traveling salesman problem and minimum spanning trees: part II," Mathematical Programming 1 (1971), 6-25.
24. M. Held and R. M. Karp and P. Wolfe, "Large Scale Optimization and the Relaxation Methods," proceedings of 25th Annual ACM Conference, August 1972, Boston, 507-509.
25. H. W. Kuhn, "Approximate Search for Fixed Points," pp. 199-211 in Computing Methods in Optimization Problems - 2, edited by L. A. Zadeh, L. W. Neustadt, A. V. Balakrishnan, Academic Press, 1969.
26. L. S. Lasdon, "Duality and decomposition in mathematical programming," IEEE Trans. System Sci. and Cybernetics 4 (1968), 86-100.
27. T. Motzkin and I. J. Schoenberg, "The relaxation methods for linear inequalities," Canadian Journal of Mathematics 6 (1954), 393-404.
28. G. L. Nemhauser and W. B. Widhelm, "A Modified Linear Program for Columnar Methods in Mathematical Programming," Operations Research 19 (1971), 1051-1060.
29. R. T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, N. J., 1970.



30. H. Scarf, "The Approximation of Fixed Points of a Continuous Mapping," SIAM Journal of Applied Mathematics 15 (1967), 1328-1343.
31. M. Simonard, "Linear Programming," Prentice-Hall, Englewood Cliffs, N. J., 1966.
32. H. M. Wagner and T. M. Whitin, "Dynamic Version of the Economic Lot Size Model," Management Science 5 (1958), 89-96.
33. M. H. Wagner, "Constructive Fixed Point Theory and Duality in Nonlinear Programming," Technical Report No. 67, Operations Research Center, Massachusetts Institute of Technology, Cambridge, Mass., September 1971.
34. W. I. Zangwill, Nonlinear Programming: A Unified Approach, Prentice-Hall, Englewood Cliffs, N. J., 1969.





BASIS

Date Due

DEC 01 72

MAY

JUN 7 73

MAY 07

MAY 07 81

APR 01 1991

Lib-26-67

T-J5 143 w no.624- 72  
Fry, Ronald Eu/The accelerated summer  
636659 D\*BKS 00028904



3 9080 000 761 681

MIT LIBRARIES



3 9080 003 670 897

HD28.M414 no.626- 72a  
Zannetos, Zeno/The economic impact of  
637016 D\*BKS 00027782



3 9080 000 748 357

MIT LIBRARIES



3 9080 004 667 322

HD28.M414 no.629-72  
Sertel, Murat /The fundamental continu  
637018 D\*BKS 00027759



3 9080 000 747 755

MIT LIBRARIES



3 9080 003 701 940

MIT LIBRARIES



3 9080 003 670 921

MIT LIBRARIES



3 9080 003 671 911

MIT LIBRARIES

DUPL



3 9080 003 701 965

MIT LIBRARIES

DUPL



3 9080 003 670 996

